

TNO Defence Research

TNO Physics and Electronics
Laboratory

TD 92-3314

Phone +31 70 328 42 21

TNO-report
FEL-92-B356

copy no.

title

Applications of motion estimation in image processing

AD-A263 282



DTIC
ELECTE
APR 28 1993
S C D

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

98 4 26 11 8

93-08819



TNO Defence Research

TNO Physics and Electronics Laboratory

Oude Waalsdorperweg 63
2597 AK The Hague
P.O. Box 96664
2509 JG The Hague
The Netherlands

Fax +31 70 328 08 61
Phone +31 70 328 42 21

TNO-report
FEL-92-B356

copy no.

88

title

Applications of motion estimation in image processing

author(s):

W.J.C. Beck

date:

October 1992

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

classification

title : unclassified

abstract : unclassified

report text : unclassified

All rights reserved.

No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

In case this report was drafted on instructions, the rights and obligations of contracting parties are subject to either the 'Standard Conditions for Research Instructions given to TNO', or the relevant agreement concluded between the contracting parties.

Submitting the report for inspection to parties who have a direct interest is permitted.

TNO

no. of copies : 40

no. of pages : 18 (excluding RDP and distribution list)

no. of appendices : -

All information which is classified according to Dutch regulations shall be treated by the recipient in the same way as classified information of corresponding value in his own country. No part of this information will be disclosed to any party.

Netherlands organization for applied scientific research

TNO Defence Research consists of the TNO Physics and Electronics Laboratory, the TNO Chemistry Laboratory and the TNO Biology Laboratory.



The Standard Conditions for Research Instructions given to TNO, as filed at the Registry of the District Court and the Chamber of Commerce in The Hague shall apply to all instructions given to TNO.

report no. : FEL-92-B356
title : Applications of motion estimation in
image sequence processing
author(s) : Wouter Beck
institute : TNO Physics and Electronics Laboratory

date : October 1992
NDRO no. : -
no. in pow '92 : 718.3

Research supervised by : G.W.M. v. Mierlo
Research carried out by : W.J.C. Beck, S.C. v.d. Marck

ABSTRACT (UNCLASSIFIED)

This report presents examples of image sequence processing using motion estimation. We present results of noise reduction by motion compensated temporal filtering in a noisy image sequence and we give examples of the detection of moving targets in an air-to-ground IR image sequence.

rapport no. : FEL-92-B356
titel : Toepassingen van bewegingsschatting in beeldbewerking
op beeldreeksen
auteur(s) : Wouter Beck
instituut : Fysisch en Elektronisch Laboratorium TNO
datum : October 1992
hdo-opdr.no. : -
no. in iwp '92 : 718.3

Onderzoek uitgevoerd o.l.v. : G.W.M. v. Mierlo

Onderzoek uitgevoerd door : W.J.C. Beck, S.C. v.d. Marck

SAMENVATTING (ONGERUBRICEEERD)

In dit rapport worden voorbeelden gegeven van beeldbewerking op beeldreeksen waarbij gebruik wordt gemaakt van bewegingsschatting. Er worden resultaten getoond van ruisreductie door bewegingsgecompenseerde temporele filtering en van de detectie van bewegende doelen in een *air-to-ground* IR beeldreeks.

ABSTRACT	2
SAMENVATTING	3
CONTENTS	4
1 INTRODUCTION	5
2 NOISE REDUCTION IN IMAGE SEQUENCES	6
2.1 Filtering along motion trajectories	6
2.2 Examples	7
3 DETECTION OF MOVING TARGETS	10
3.1 Principle of detection	10
3.2 Example of target detection using target motion	14
REFERENCES	18

INTRODUCTION

Motion perception is an important source of information for the human visual system. The determination of our motion relative to the environment as well as the determination of the three dimensional structure of the environment largely depend on the interpretation of visual motion. The human visual system is capable of extracting information from a sequence of images that is hard to extract from the individual images. An example is the interpretation of a very noisy image sequence. By using spatial and temporal correlation we are able to "see through the noise." Sometimes, visual detection of an object fully depends on the perception of motion. This is illustrated by the ease with which we see an otherwise successfully camouflaged object as soon as it moves.

The apparent motion of brightness patterns observed when a camera is moving relative to the objects being imaged is called *optical flow*. Optical flow can be represented by a two dimensional vector field. Loosely speaking, the optical flow field links a pixel at the position (x, y) to the corresponding pixel at position $(x + u(x, y), y + v(x, y))$ in the next image. Ideally, both pixels correspond to the same physical object point in the scene. In practice, this is hard to achieve because there is an infinite number of vector fields that is consistent with the data. Possible approaches to the problem of estimating the optical flow field are described in a separate report [2]. This report also describes the particular approach to the motion estimation problem we have taken. See also [1].

The goal of this report is to give examples of image sequence processing using motion estimation. All examples were obtained by simple and straightforward methods. The examples show what kind of improvements or detection results can be expected, rather than being optimal results in some sense.

Chapter 2 of this report is concerned with noise reduction in a noisy image sequence. It presents some results on IR image sequences.

Chapter 3 deals with the detection of moving objects in image sequences. Typically, these sequences are recorded from a moving platform. The aim is to detect objects moving relative to the moving background.

There are numerous other applications of image motion estimation. These include image sequence coding, satellite image processing, medical image processing, robot vision, obstacle avoidance, image sequence stabilization etc.

2 NOISE REDUCTION IN IMAGE SEQUENCES

This chapter is concerned with noise reduction in image sequences by motion compensated temporal filtering. We present some preliminary results obtained with straightforward methods.

2.1 Filtering along motion trajectories

By $I(\mathbf{x}, t)$ we denote the image brightness function, where vector \mathbf{x} denotes the spatial coordinates and t denotes time. Let $\mathbf{v}(\mathbf{x}, t)$ be the displacement of the image point at (\mathbf{x}, t) between time $t - \Delta t$ and t , where Δt denotes the temporal sampling interval. Assuming that image brightness for an object point is conserved over time, we can write

$$I(\mathbf{x}, t) = I(\mathbf{x} - \mathbf{v}(\mathbf{x}, t), t - \Delta t) \quad (2.1)$$

Obviously, \mathbf{v} is undefined when an object is occluded or when it is newly exposed. In general \mathbf{v} will be a slowly varying function of the spatial coordinates with discontinuities at the edges of moving objects. A spatiotemporal volume can be formed by stacking the consecutive frames of the sequence. A physical point in the scene traces out a trajectory in this spatiotemporal volume during the time it is visible in the sequence. The brightness value along this trajectory forms a one dimensional signal. This signal is assumed to consist of a deterministic image component and an additive noise component. Variation of the image component is due to change in the luminance of the object. This variation is assumed to be relatively slow, so that the image component is a low bandwidth signal. The additive noise is assumed to be uncorrelated with the image signal. Lowpass filtering along the motion trajectory can significantly reduce the noise component. The filter operation along the motion trajectory can be either linear or non-linear. When the image noise is additive Gaussian noise, independent in each pixel and of fixed variance along a motion trajectory, then it can be shown that the sample mean along a motion trajectory is the maximum likelihood estimator for the grey value of the pixel. In this case, the linear estimator will yield the best signal to noise ratio in the result. On the other hand, a non-linear filter may be more robust to errors in the displacement estimate and the non-validity of the noise model e.g. in the case of dead pixels in the images. In addition, a non-linear filter might be able to deal with occlusion and exposure effects more adequately. The choice of filter will generally depend on the ease of implementation and the particular distortions in the image sequence. The number of frame stores can be reduced if the used filter is recursive [5].

With regard to exposure and occlusion effects, it would be of interest to know exactly the lifetime of a motion trajectory. Unfortunately this is a hard problem. It requires the identification of image areas that are newly exposed and image areas that are just occluded in each frame of the image sequence. Most current motion estimators are not able to solve this problem reliably.

2.2 Examples

This section presents examples from a IR image sequence. Figure 2.1 shows the reference image from this sequence. All computations are performed relative to this image. Figure 2.1 shows the average of six motion compensated images, 3 forward in time and 3 backward in time, and the reference image. This result is an example of a linear filtering operation along the motion trajectories. *Figure 2.2 shows the median of six motion compensated images and the reference image. This result is an example of a non-linear filtering operation along the motion trajectories. It is clear from both the linear and the non-linear results that the noise is significantly reduced without affecting the sharpness. In this case, the linear filter yields a visually more pleasing result. In both results details become visible that are hard to infer from a single noise corrupted image. Figure 2.2 shows the result of running the contrast enhancement algorithm of [4] on the image of figure 2.1. This algorithm locally adjusts the contrast. Without the noise reduction furnished by the motion compensated filtering, local contrast enhancement merely amplifies the noise and other image distortions.*



Figure 2.1: (Top) Original reference image from an IR image sequence. (Bottom) Mean along motion trajectories of six motion compensated images and the reference image.

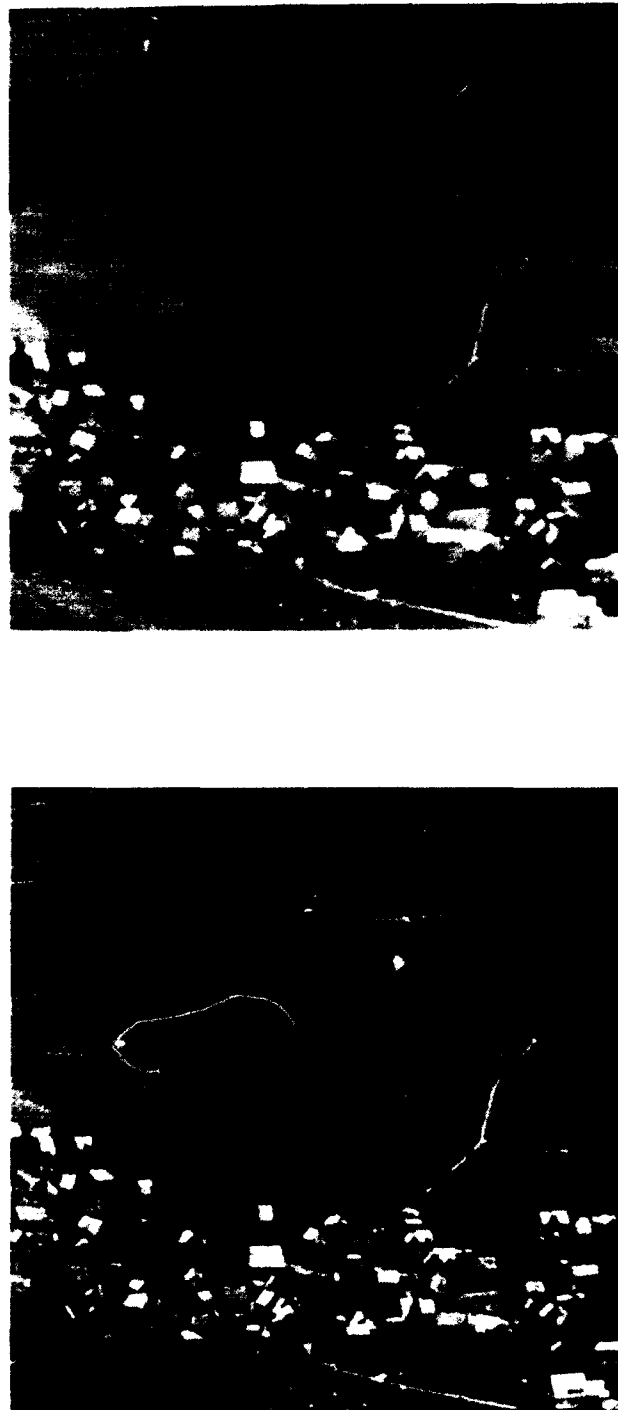


Figure 2.2: (Top) Median along motion trajectories of six motion compensated images and the reference image. (Bottom) Contrast enhanced version of bottom image of figure 2.1

3 DETECTION OF MOVING TARGETS

This chapter is concerned with the detection of moving targets from an image sequence produced by a moving camera. The proposed approach is particularly well suited for air-to-ground applications. In principle, it is possible to detect camouflaged targets moving relative to a textured background.

3.1 Principle of detection

Algorithms for the detection of dim, low contrast targets usually consist of two stages. Firstly, the algorithm selects a number of potential targets, for example bright spots. Due to clutter¹ this usually results in a large number of false alarms. The second stage therefore has to reject the falsely selected objects. This can be done by combining information over frames, or by use of contextual information.

In this report, we choose to detect targets on basis of their motion. Our approach consists of essentially two stages:

1. motion estimation,
2. target detection in the motion compensated image sequence.

We assume we have to deal with an essentially stationary scene that is being imaged from a moving platform (e.g. helicopter). If we are able to estimate a sufficiently accurate 2-D vector field that maps one frame in the sequence to the next, we can in principle predict one frame from the previous one. The principle to detect moving targets is particularly simple and amounts to analyzing the image sequence on the occurrence of unexpected events. In this context, unexpected events are temporal variations of the image brightness function that are impossible to predict and that can not be accounted for by noise. Thus, targets are detected by analyzing the difference between the predicted and the actual image.

As a simple example of this detection principle, consider a stationary camera imaging a stationary scene. From one or a collection of images (frames) it is possible to predict the next frame to be acquired by the camera. The prediction is simply the previous frame or an estimate based on a collection of previous frames, such as the pixel-wise mean. The only difference between the predicted

¹ In this report, clutter is loosely defined as the amount of target-like objects in a scene.

frame and the actually acquired frame will be due to noise that is introduced somewhere in the imaging process. When an object in the scene moves, it is generally impossible to predict the next frame from a limited history of past frames. This is due to the fact that the object uncovers background that is not visible in previous frames and hence hard to predict. Even when motion estimation is used, and the moving part of the scene is warped², it is impossible to predict the background that is uncovered by the moving object. Subtracting the predicted frame from the actual frame will generally produce large differences at image locations corresponding to covered and uncovered background. Automatic detection based on this principle is feasible by comparing the differences with the noise statistics of the difference image.

A similar detection principle can be used when a scene is imaged from a moving platform. In this case we have to perform image registration prior to subtraction. Image registration is done by estimating the image motion field and warping the first frame to the next accordingly. The image motion is estimated using the method described in [1, 2]. Image warping generally involves interpolation. In the examples shown in this chapter we used *bi-linear interpolation*.

In evaluating the difference images thus obtained, we have to distinguish several possibilities.

1. When the image motion estimate is perfect, the scene has no appreciable depth discontinuities, and there are no moving objects in the scene, we expect the difference image to be a sample from a 2-D random noise process. The noise is a mixture of image noise and noise due to the interpolation process.
2. When the motion estimate is accurate, there are no moving objects, but there is considerable depth variation in the scene, we expect uncovered background adjacent to physical edges of foreground objects. This is the parallax effect. Generally, this will result in large amplitudes in the difference image at locations corresponding to covered and uncovered background, while the rest of the difference image is characterized by random noise. In air-to-ground imagery the large response areas will usually be chain-like, for example the outline of a hill. Although the large response areas will generally not correspond to moving targets, they are nonetheless of interest because they often correspond to previously unexposed parts of the scene. In some applications it may be of interest to perform extra processing on parts of the scene that are newly exposed.
3. When there are moving objects in the scene and the motion estimate is such that this object motion is correctly captured, the difference image will show large amplitudes at locations of

² In this context, *warping* is the process in which one image is registered with another and resampled (interpolated) at the grid positions of the reference image.

covered and un-covered background if the background is sufficiently textured. This enables us to detect camouflaged objects.

4. When there are small, moving objects we may be unable to capture object motion correctly. This behaviour may be forced by only using large scale image structure in the motion estimate. In this case the difference image will generally display a small area with a large positive response adjacent to a small area with a large negative response. This case is of practical interest in target acquisition applications at long stand off ranges.

To automate the detection process, we have to make a number of assumptions about the image noise statistics. In the examples shown in this report, the noise was assumed to be additive zero-mean Gaussian noise, independent in each pixel. The noise statistics are obtained from a fit of a Gaussian to the sample histogram of the difference image. This is more robust than calculating the usual sample statistics because the influence of the outliers (targets!) is reduced. From the standard deviation thus obtained, a statistically meaningful threshold may be obtained.

The confidence in the presence of potential targets may be increased by correlating the detection results over time. This may involve more or less sophisticated techniques such as [3].

The overall procedure for detection of moving targets can be summarized by the scheme of figure 3.1

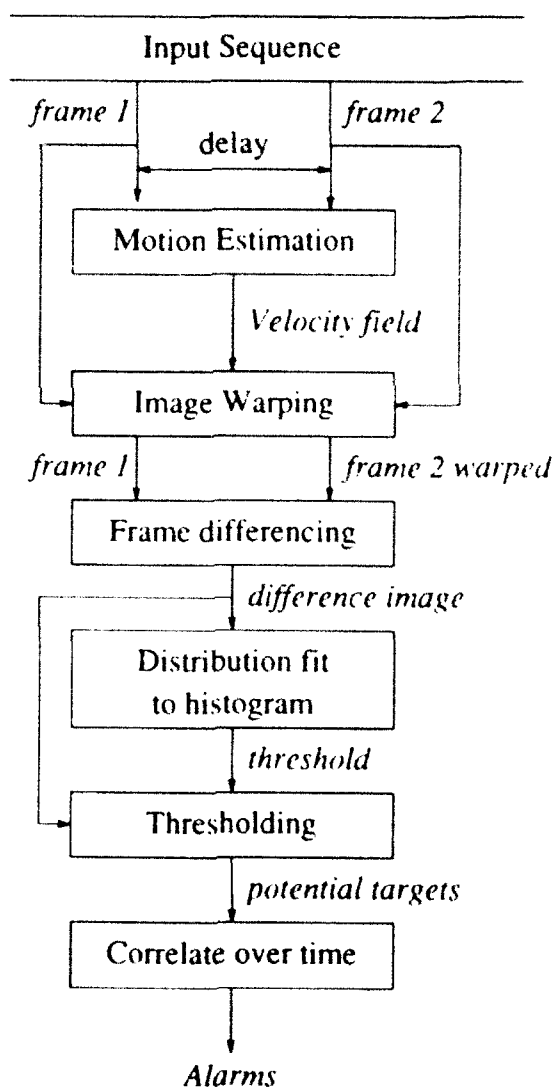


Figure 3.1: Block diagram for the detection of moving targets in an image sequence obtained by a moving camera.

3.2 Example of target detection using target motion

The example shown in figure 3.4 is an example of case 4 in the enumeration of pages 11 and 12. The upper photograph of figure 3.4 shows a frame from an air-to-ground IR image sequence. In this sequence there are several moving targets, cars on the roads. Notice that in this sequence the contrast is inversed, i.e. hot areas appear dark. For the present algorithm, this makes no difference.

For the target detection we used three frames $f(t)$ at times t_- , t_0 and t_+ . The image motion between $f(t_0)$ and $f(t_-)$ and between $f(t_0)$ and $f(t_+)$ was estimated using the phase based estimator described in [1, 2]. Because this image sequence is contaminated by a fair amount of noise and sensor artefacts, and because this image sequence lacks small scale image structure in certain parts of the scene, we used a planar patch model to improve the estimated image motion.

It can be shown [6] that the planar patch model is described by the mapping:

$$X' = \frac{A_{11}X + A_{12}Y + A_{13}}{A_{31}X + A_{32}Y + 1}, \quad (3.1)$$

$$Y' = \frac{A_{21}X + A_{22}Y + A_{23}}{A_{31}X + A_{32}Y + 1}. \quad (3.2)$$

Equations (3.1) and (3.2) define a mapping from the two-dimensional image-space (X, Y) at time $t = t_1$ onto the image-space (X', Y') at $t = t_2$, see figure 3.2. The eight non-trivial parameters A_i are the so called *pure parameters* ($A_{ii} = 1$). They are uniquely determined for a given motion and planar patch. The pure parameters are estimated from the image motion vectors produced by the phase based motion estimator. Both $f(t_-)$ and $f(t_+)$ were warped according to the estimated model (3.1) and (3.2) to obtain estimates valid at $t = t_0$. These warped images are denoted by $f(t_- \rightarrow t_0)$ and $f(t_+ \rightarrow t_0)$.

First, we form the difference images $d_- = f(t_0) - f(t_- \rightarrow t_0)$ and $d_+ = f(t_0) - f(t_+ \rightarrow t_0)$. Figure 3.3 shows a histogram of d_+ . From figure 3.3 it is clear that this distribution is very well approximated by a Gaussian distribution, as shown by the dashed line. The parameters of this Gaussian were determined using a non-linear least squares fit to the histogram.

Next, we apply a thresholding procedure to the difference images d_- and d_+ . A threshold factor θ is selected. Let $p(x, y)$ be the pixel value at location (x, y) in either d_- or d_+ , and let μ and σ be the corresponding mean and standard deviation, respectively, as determined by the histogram fit. We define s by

$$s(x, y) = \frac{p(x, y) - \mu}{\sigma}. \quad (3.3)$$

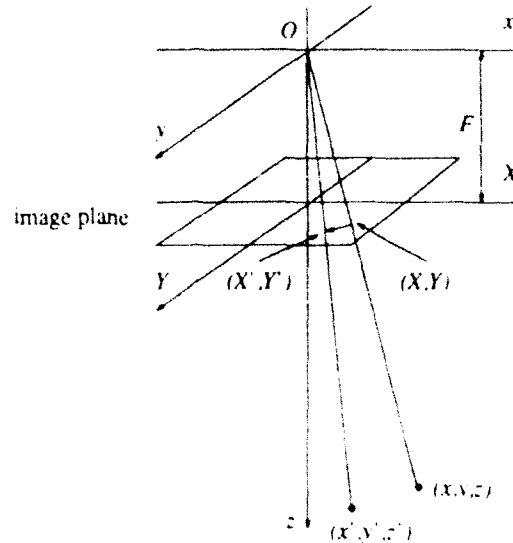


Figure 3.2: Basic geometry for three-dimensional motion estimation. Lower case letters refer to 3-D scene coordinates, whereas the image coordinates are denoted by capitals. F is the focal length of the camera. (x, y, z) denote the coordinates of an object at $t = t_1$, (x', y', z') give the the coordinates of the same object at $t = t_1$.

The result of thresholding procedure is determined by:

$$o(x, y) = \begin{cases} 0 & \text{if } |s(x, y)| < \theta/2 \\ \frac{2s(x, y)}{\theta} - \text{sign}(s(x, y)) & \text{if } \theta/2 < |s(x, y)| < \theta \\ \text{sign}(s(x, y)) & \text{if } |s(x, y)| > \theta \end{cases} \quad (3.4)$$

where $\text{sign}(\xi)$ is defined by

$$\text{sign}(\xi) = \begin{cases} -1 & \text{if } \xi < 0 \\ 0 & \text{if } \xi = 0 \\ 1 & \text{if } \xi > 0 \end{cases} \quad (3.5)$$

This yields two frames of which almost all pixels are zero except for a number of positive and negative 'blobs' with values between 0 and 1 and -1 and 0, respectively. This thresholding procedure has the advantage that it retains target responses that are not very strong. Of course, these 'weak' target responses have to be confirmed later on. Next, we discard all non-zero pixels in both frames that have opposite signs at corresponding positions. These 'cleaned' images are referred to as c_- and c_+ .

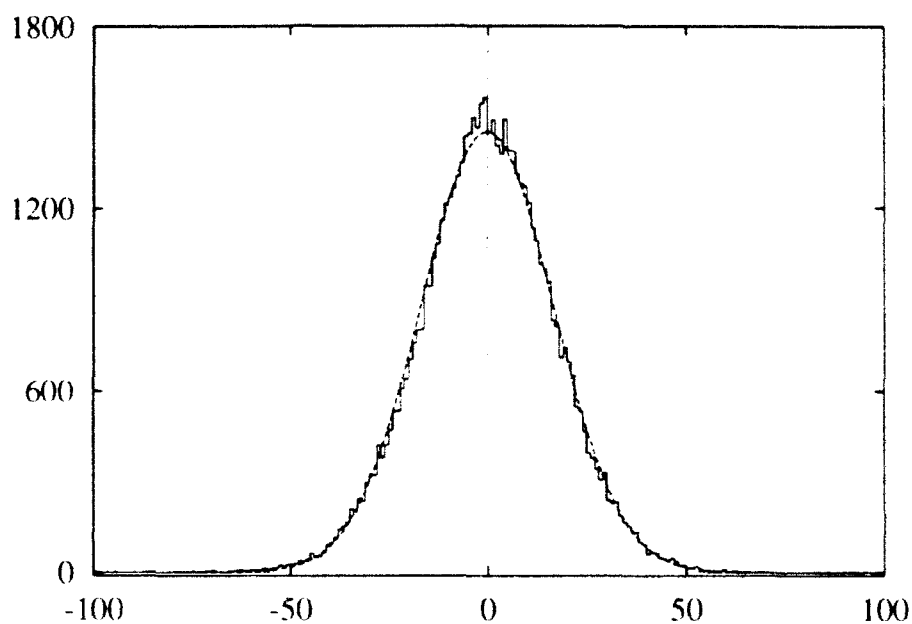


Figure 3.3: Histogram of the difference image d_+ , obtained by subtracting the motion compensated image $f(t_+)$ from the reference image $f(t_0)$. The dashed line represents the fitted Gaussian.

In the next step, we combine the images c_- and c_+ by pixel-wise multiplication. The positive blobs in this image, denoted by T_0 , correspond to potential targets.

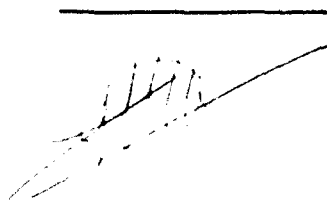
The detection result shown in figure 3.4 was obtained by requiring that each target in T_0 corresponds to an object in image $f(t_0)$ that significantly differs in greyvalue from its neighborhood as determined by an inverse median filter. Inverse median filtering amounts to forming the difference between the original image and a median filtered version of it.



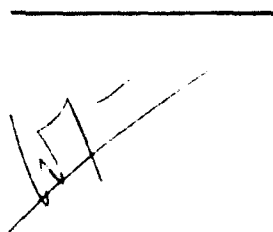
Figure 3.4: (Top) One image from an air-to-ground IR image sequence. (Bottom) Four consecutive frames with the detected moving targets in red. There are a few false alarms in individual frames. However, only the true targets are consistently detected. The false alarms could be eliminated by requiring consistency over time.

REFERENCES

- [1] W.J.C. Beck.
Hierarchical computation of image velocity from local phase information.
In *Proc. 11th IAPR Int. Conf. on Pattern Recognition, The Hague*, 1992.
- [2] W.J.C. Beck.
Phase based image motion estimation (fel-92-b302).
Technical report, TNO Physics and Electronics Laboratory, 1992.
(NATO CONFIDENTIAL).
- [3] S.D. Blostein and T.S. Huang.
Detecting small, moving objects in image sequences using sequential hypothesis testing, part i:
Algorithms and analysis.
IEEE Trans. on Acoustics Speech and Signal Processing, 1990.
- [4] F.P.H. de Vries.
Automatic, adaptive, brightness independent contrast enhancement.
Signal Processing, 21:169-182, 1990.
- [5] E. Dubois and S. Sabri.
Noise reduction in image sequences using motion compensated temporal filtering.
IEEE Trans. on Communications, COM-32:826-831, 1984.
- [6] R.Y. Tsai and T.S. Huang.
Estimating three-dimensional motion parameters of a rigid planar patch.
IEEE Trans. on Acoustics Speech and Signal Processing, 29:1147-1152, 1981.



G.W.M. van Mierlo
(group leader)



W.J.C. Beck
(author/project leader)

UNCLASSIFIED

REPORT DOCUMENTATION PAGE

(MOD-NL)

1. DEFENSE REPORT NUMBER (MOD-NL) TD92-3314	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER FEL-92-B356
4. PROJECT/TASK/WORK UNIT NO. 20430	5. CONTRACT NUMBER -	6. REPORT DATE OCTOBER 1992
7. NUMBER OF PAGES 18 (EXCL. RDP & DISTRIBUTION LIST)	8. NUMBER OF REFERENCES 6	9. TYPE OF REPORT AND DATES COVERED FINAL
10. TITLE AND SUBTITLE APPLICATIONS OF MOTION ESTIMATION IN IMAGE SEQUENCE PROCESSING		
11. AUTHOR(S) W.J.C. BECK		
12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TNO PHYSICS AND ELECTRONICS LABORATORY, P.O. BOX 96864 2509 JG THE HAGUE OUDE WAALSDORPERWEG 63, THE HAGUE, THE NETHERLANDS		
13. SPONSORING/MONITORING AGENCY NAME(S) TNO PHYSICS AND ELECTRONICS LABORATORY, THE HAGUE, THE NETHERLANDS		
14. SUPPLEMENTARY NOTES		
15. ABSTRACT (MAXIMUM 200 WORDS, 1044 POSITIONS) THIS REPORT PRESENTS EXAMPLES OF IMAGE SEQUENCE PROCESSING USING MOTION ESTIMATION. WE PRESENT RESULTS OF NOISE REDUCTION BY MOTION COMPENSATED TEMPORAL FILTERING IN A NOISY IMAGE SEQUENCE AND WE GIVE EXAMPLES OF THE DETECTION OF MOVING TARGETS IN AN AIR-TO-GROUND IR IMAGE SEQUENCE		
16. DESCRIPTORS TARGET DETECTION NOISE REDUCTION IMAGE MOTION COMPENSATION INFRARED IMAGES		IDENTIFIERS IMAGE PROCESSING IMAGE SEQUENCES MOTION ESTIMATION
17a. SECURITY CLASSIFICATION (OF REPORT) UNCLASSIFIED	17b. SECURITY CLASSIFICATION (OF PAGE) UNCLASSIFIED	17c. SECURITY CLASSIFICATION (OF ABSTRACT) UNCLASSIFIED
18. DISTRIBUTION/AVAILABILITY STATEMENT UNLIMITED		17d. SECURITY CLASSIFICATION (OF TITLES) UNCLASSIFIED